DOCUMENT RESUME

ED 290 803 TM 870 758

AUTHOR Becker, Betsy Jane
TITLE An Evaluation of Approximations to the Distribution of the Sample Significance Value.
PUB DATE Apr 87
NOTE 33p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, DC, April 20-24, 1987).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Effect Size; *Meta Analysis; Probability; *Sample Size; *Statistical Distributions; *Statistical Significance; *Statistical Studies
IDENTIFIERS Asymptotic Distributions

ABSTRACT

The random variable p and its functions figure in several "tests of combined significance," meta-analysis summaries based on sample significance values, and ps have been used singly, as well as in other tests for evaluating the outcomes of individual research studies. In this work, asymptotic distributions of the sample one-sided significance value, p, and several functions of p were examined. Asymptotic normal distributions of p, of the standard normal deviate associated with the one-sided p value, z(p), and of the natural logarithm of the significance value, log(p), were studied. Cumulative normal distribution functions of the transformed variables were calculated and compared to the computed exact distributions of ps associated with one-sample t tests. Distributions based on the z(p) transformation appeared quite accurate even for very small sample sizes, while others were quite inaccurate. One of the distributions based on z(p) was particularly accurate and can be applied to ps not based on t tests. However, the standard normal approximation which is much simpler, was accurate for small effect sizes. (Author/JGL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED290803

An Evaluation of Approximations to the Distribution of the Sample Significance Value

Betsy Jane Becker

Michigan State University

April, 1987

Paper presented at the 1987 meeting of the American Educational Research Association, Washington, D.C.

Running head: APPROXIMATE DISTRIBUTIONS

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Betsy Jane Becker

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 870 852

An Evaluation of Approximations to the Distribution of the Sample Significance Value

Abstract

The random variable *p* and its functions figure in several "tests of combined significance," meta-analysis summaries based on sample significance values, and *p*s have been used singly, as well as in other tests for evaluating the outcomes of individual research studies. In this work, asymptotic distributions of the sample one-sided significance value, *p*, and several functions of *p* were examined. Asymptotic normal distributions of *p*, of the standard normal deviate associated with the one-sided *p* value, *z*(*p*), and of the natural logarithm of the significance value, log(*p*), were studied. Cumulative normal distribution functions of the transformed variables were calculated and compared to the computed exact distributions of *p*s associated with one-sample *t* tests. Distributions based on the *z*(*p*) transformation appeared quite accurate even for very small sample sizes, while others were quite inaccurate. Implications for the use of the approximations conclude the paper.

An Evaluation of Approximations to the Distribution of the Sample Significance Value

Until recently social scientists have paid little attention to the statistical properties of the significance value or *p* value. Though probabilities figure in every statistical hypothesis test conducted, they have often been treated as the static final outcomes of statistical tests, indicating little more than rejection or acceptance of null hypotheses.

With the recent increasing interest in meta-analysis, the quantitative synthesis of research results (Glass, 1976), interest has increased in the use of obtained probabilities to represent study outcomes. Many procedures have been suggested to that end, for example, several procedures use the inverse normal transformation of *p* (i.e., $z(p)$), to provide planned contrasts among study results (Rosenthal & Rubin, 1979; Strube, 1985). Because such statistics are functions of significance values, their distributions depend on distributions of functions of the sample significance value.

In this paper several asymptotic distributions of functions of *p* are studied. These distributions are meant to be applied in situations in which the null hypothesis is false, since distributions of many functions of *p* are known under the null hypothesis. The distributions studied here were drawn from the literatures of statistics and methodology for the social sciences.

The first section of the paper is a brief review of the uses suggested for significance levels and functions of significance levels. Section two introduces notation and a model for the results of a study employing a *t* test of the single parameter δ, and also introduces the functions of *p* to be studied. Section three begins with a discussion of previous work on

approximations to distributions of functions of the significance value, and presents the distributions to be examined. The usefulness of obtaining normal approximations to the nonnull distribution of *p* and functions of *p* is discussed. Section four reports the results of numerical comparisons of the asymptotic cumulative distributions to the (computed) exact cumulative distribution of *p*'s from *t* tests. The last section presents a summary of the results, recommendations concerning use of the approximations, and comments on the implications of the results for the suggested uses of significance values.

Uses of Significance Values

The primary impetus for studying distributions of functions of significance values comes from the use of *p* values in meta-analysis. The more familiar approach to meta-analysis involves accumulating and analyzing measures of effect magnitude from series of studies.

However, another approach to quantitative reviewing involves calculating an "overall level of significance" (Rosenthal, 1978, p. 185) using the sample significance values from the individual studies. This overall probability for the series of studies is based on one of a set of tests of combined significance. Rosenthal (1984) and Cooper and Rosenthal (1980) most recently brought these tests to the attention of social scientists and their use has grown. The tests are calculated using one-sided obtained probabilities from independent hypothesis tests.

Another related use suggested for significance values in meta-analysis has been as a means for comparing results of particular studies (Rosenthal & Rubin, 1979; Strube, 1985). Rosenthal labels these as "diffuse" and "focused" tests for comparing studies "in terms of significance testing"

(1983, p. 5). The tests resemble planned and post hoc comparisons typically applied in analysis of variance, but are based on the inverse-normal-transformed sample probabilities (the z(p) values) from the studies.

Notation for Study Results

Consider a study examining the parameter δ via a one-sample design. Let δ represent, for simplicity, a standardized mean parameter. A common example of a study using this kind of design is the one-group pretest-posttest study (Campbell & Stanley, 1963). In such a study the observation Y_j would represent the difference between the posttest and pretest scores, respectively, of the jth individual.

Denote the sample size as n and assume that the scores Y_j are normally distributed with mean μ and variance σ^2. Thus

$$Y_j \sim N(\mu, \sigma^2), \quad \text{for } j = 1, \ldots, n. \tag{1}$$

The term effect size has been used in statistics to label many effect-magnitude measures, here the (population) effect size is the standardized mean, $\delta = \mu/\sigma$. The sample effect-size estimate is $g = \overline{Y}/S$, where $\overline{Y}$ is the sample mean and S is the usual standard-deviation estimate.

The k parameters $\delta_1, \ldots, \delta_k$ are the population values about which inferences are made when series of k sample standardized means are analyzed.

The Significance Value

The relationship between the effect size and the sample significance value (or p value) arises through the relation of both to the sample t test. The standardized mean g is a simple linear transformation of the one-sample t test about the parameter δ. The relationship of the statistic g to the

test statistic t is shown as

$$t_0 = \frac{\bar{Y} - \mu}{S / \sqrt{n}} = g\sqrt{n}. \qquad (2)$$

This t statistic has $n - 1$ degrees of freedom.

The one-sided (directional) significance value that represents the standardized-mean estimate g is the upper-tail probability from a t test of the null hypothesis of a zero population mean, a zero effect size, or equivalently, a zero noncentrality parameter. The null model is

$$H_0: \ \mu = 0 \quad \text{or} \quad \delta = 0 \quad \text{or} \quad \sqrt{n}\,\delta = 0, \qquad (3)$$

versus the alternative that the population mean, effect size, or noncentrality parameter exceeds zero. The relationship of the population mean tested in H_0 to the population effect-size parameter is clear. The effect size is simply the mean divided by the standard deviation of scores. Thus the t test that the population mean is positive is also a test for a positive population effect size or a positive noncentrality parameter. The directional alternatives can be written symbolically as

$$H_1: \ \mu > 0 \quad \text{or} \quad \delta > 0 \quad \text{or} \quad \sqrt{n}\,\delta > 0.$$

Hypotheses about nonzero values for μ (such as μ_0) can be accounted for in this notational model by defining a value μ' as $\mu' = \mu - \mu_0$ and testing $H_0: \mu' = 0$.

The p value associated with the t test thus represents the same substantive effect as the sample effect size. The one-sided sample

significance value is the upper-tail p value (p_0) associated with the sample value t_0 given in Equation 2. The significance value p_0 is formally defined in terms of g via

$$p_0 = \int_{t_0}^{\infty} f(t, m)\, dt, \tag{4}$$

where $t_0 = g\sqrt{n}$, and $f(t, m)$ is the probability integral function for the central t distribution with $m = n - 1$ degrees of freedom. Another way of denoting p_0 is to refer to the cumulative probability distribution function F (here, for the t distribution). Since $F(t_0)$ is the probability of finding a value equal to or less than t_0, $p_0 = 1 - F(t_0)$.

Similar relationships hold for other parameters and the estimators and probabilities representing those parameters (e.g., for effect sizes and ps from two-sample t tests). However, different models for study results would apply to these other situations.

Distributions of the Significance Value and Related Functions

The distribution of any significance value under the null hypothesis is uniform on the interval between zero and one, regardless of the test from which it is obtained. Independent significance values from series of tests thus are identically distributed when the null hypotheses underlying all of the tests are true, again regardless of the kinds of statistical tests used.

The exact nonnull distribution of the significance value is much more complex than the null distribution, and differs according to the type of statistical test used. E. S. Pearson (1938) and others have shown that the general form of the density function f_p of a p value can be written as

$$f_p(p) = \left. \frac{f(t|H_1)}{f(t|H_0)} \right|_{t = F^{-1}(1 - p)}$$

This is the ratio of the true nonnull density of the statistic t (that is, the density under the model specified by H_1) to the null or central density of t, where t is considered as a function of p (i.e., $t = F^{-1}(1 - p)$). Even for the simple example of the one-sample t statistic this density is complex. If the effect size from a one-sample t takes on the value δ_1 under H_1, the formula for $f_p(p)$, the noncentral density of the significance value, is

$$\frac{\Gamma(1/2)\ \exp(-\delta_1^2/2)}{\Gamma(n/2)} \sum_{s=0}^{\infty} \Gamma([n + s]/2)\ \frac{\delta_1^s}{s!}\ (2t^2/[n - 1 + t^2])^{s/2}. \quad (5)$$

evaluated at $t = F^{-1}(1 - p)$, based on a sample of size n. This particular density function illustrates two features of all exact alternative-hypothesis densities of significance values. Such functions depend on the value of the effect-size parameter under the alternative model, and they depend on the sample size.

One reason that statisticians have not relied solely on analyses and summaries of untransformed p values is that the behavior of the significance level is influenced by the sample size. It is well known that significance tests become large due to the use of larger samples as well as from finding larger effects. Thus direct "comparisons of p values have little meaning" (Gibbons & Pratt, 1975, p. 21) when the null hypothesis is false.

Many statisticians have considered summaries of transformed p values, which are also identically distributed under the null hypothesis. Most

tests of combined significance and other comparisons of study results involve statistics which are functions of transformed significance values. The distributions of these statistics depend fundamentally on the distribution of the individual significance value. Exact densities of the statistics under alternative hypotheses are very complicated because they involve convolutions of densities of the form shown in Formula 5.

Asymptotic distributions are useful because they are often simpler than exact (small-sample) distributions. Asymptotic normal distributions are particularly useful since the standard normal distribution is familiar and easily evaluated. Precise computational algorithms to evaluate the standard normal distribution are available, and standard normal tables are provided in most elementary statistics books as well.

Asymptotically normal alternative-hypothesis distributions have been proposed for two functions of the significance value: the logarithmically transformed p value, $\log(p)$, and the p value transformed via the inverse cumulative normal distribution function, $z(p)$.

The Work of Lambert and Hall

Most of the work on distributions for p values and their functions has been done by Lambert (1978) and Lambert and Hall (1982, 1983). In her dissertation and pursuant work Lambert has studied primarily the asymptotic lognormality of p values. Most of this investigation was done in conjunction with study of the optimality properties of tests. The distributions derived by Lambert depend on the exact slope of the test associated with the probability value (see, e.g., Bahadur, 1967), which is an indicator of asymptotic efficiency. The results apply to exact p values and with only minor modification to both approximate and conditional p's.

The results are presented in Lambert and Hall (1982) for the general case of a statistic T_n and parameter θ. Here T_n represents the t test and θ is replaced by δ, the effect-size parameter for the t statistic.

Lambert's Result 1. Let T_n be a sample statistic which provides a test of H_0: $\delta = 0$, and let p be the obtained significance value associated with that test for a sample of size n. For any fixed nonnull value δ, assume that there exist constants $b(\delta)$ and $\sigma(\delta)$, with $-\infty < b(\delta) < \infty$ and $0 < \sigma(\delta) < \infty$, and such that

$$\sqrt{n}\,(T_n - b(\delta)) \sim AN(0, \sigma^2(\delta)), \tag{6}$$

as $n \to \infty$. (Often $b(\delta) = \delta$, as in the present case where $\delta = \mu/\sigma$.) With several other assumptions (detailed on their page 47) Lambert and Hall (1982) showed that p, the upper-tail p value associated with T_n, is asymptotically a lognormal variable with mean $-n\,c(\delta)$ and variance $n\,\tau^2(\delta)$. This can be written as

$$\sqrt{n}\,(-n^{-1}\log(p) - c(\delta)) \sim AN(0, \tau^2(\delta)), \tag{7}$$

as $n \to \infty$. The coefficient $c(\delta)$ in the mean is equivalent to half of the exact slope for the statistic T_n, which is the almost sure limit of $-\log(p)/n$. Formulas for $c(\delta)$ and $\tau^2(\delta)$ can be found in the statistical literature (e.g., Bahadur, 1960a, b). (Formulas for c and τ for the case of the one-sample t are given below.)

One assumption of Lambert and Hall's results is that the test statistic T_n from which the p is obtained is at least asymptotically normally

distributed. This limits the applicability of the results slightly, though in large samples the distributions of many familiar statistics (e.g., binomials, t statistics and chi squares) approach normality. One further disadvantage of the lognormal approximation is that the distribution of the function $-\log(p)/n$ is normal under alternative hypotheses, but is exponential under the null model (i.e., when $\delta = 0$).

Lambert (1978) and Lambert and Hall (1982) developed several modifications of the $-\log(p)/n$ function which have the same asymptotic distributions as $-\log(p)/n$, but which have somewhat more accurate distributions. Modifications suggested were $-\log(\sqrt{n}\,p)/n$ and $-\log(\sqrt{2\pi n}\,p)/n$, and an intermediate alternative is $-\log(2\sqrt{n}\,p)/n$. Lambert (1978, p. 65) reported that the distributions tended to be more accurate for larger effect magnitudes (i.e., the more distinct the true parameter value was from the null value which is often zero). Small-sample accuracy studies were not reported, however.

Lambert's Result 2. In her dissertation Lambert (1978) also discussed an asymptotic approximation to the distribution of the inverse-normal transformed p value, $z(p)$. Since the distribution of $z(p)$ is exactly normal under the null hypothesis, this approximation avoids the problem inherent to the distribution of $-\log(p)/n$, which has different forms under H_0 and H_1.

Define T_n, $b(\delta) = \delta$, and $\sigma^2(\delta)$ as for Model 6, and let

$$\sqrt{n}\,(T_n - \delta) \sim AN(0, \sigma^2(\delta)).$$

If p is the upper-tail p value associated with T_n, and $z(p) = -\Phi^{-1}(p)$, where Φ represents the standard normal cumulative distribution function, then

$$(z(p) - \sqrt{2\, n\, c(\delta)}) \sim AN(0, \eta^2(\delta)), \qquad (8)$$

where $\eta^2(\delta) = \tau^2(\delta)/[2c(\delta)]$, and where $c(\delta)$ and $\tau(\delta)$ are defined as for Model 7 above. Again Lambert reported no information about the accuracy of this approximation for small samples.

Approximations from Rosenthal and Rubin

A paper by Rosenthal and Rubin (1979) is the source of two other approximations to the distribution of $z(p)$. Rosenthal and Rubin discussed the distribution of $z(p)$ in the context of using comparisons of significance values as a means of comparing the results of studies. However, to avoid the problems of comparing p's directly, Rosenthal and Rubin suggest using the inverse-normal transformation to obtain the standard normal deviate associated with each p. The standard normal deviates are then compared via methods which assume that each has the standard normal distribution under the null hypothesis.

A technical discussion concluding the Rosenthal and Rubin paper gave a rationale for assuming z values to be approximately normally distributed with unit variances, and with means equal to the noncentrality parameters for the raw data ($\sqrt{n}\,\delta$ if a t test has been applied to the raw data). The distribution suggested can be formally symbolized as

$$(z(p) - \sqrt{n}\,\delta) \sim AN(0, 1). \qquad (9)$$

Rosenthal and Rubin's approximation is derived specifically for the case in which ps result from t statistics, and they do not comment on its applicability to transformed ps obtained from other statistics.

A somewhat more complicated approximation to the nonnull distribution of $z(p)$ is mentioned in Rosenthal and Rubin's technical discussion. (I label this as the Rosenthal and Rubin approximation.) This result is also derived for the case in which p's arise from t tests. Rosenthal and Rubin suggest that for large degrees of freedom f, the mean of $z(p)$ would be approximately $\sqrt{n}\,\delta\,(1 - n\delta^2/4f)$ and its variance would be approximately $[1 + (0.5 - n\,\delta^2)/f]$. This may be symbolized as

$$(z(p) - \sqrt{n}\,\delta\,(1 - n\delta^2/4f)) \sim AN(0, [1 + (0.5 - n\,\delta^2)/f]). \qquad (10)$$

This approximation was not presented as one to be applied directly, but rather as a justification for using the approximation in Model 9. To apply both this and the standard normal approximation the degrees of freedom f must be large enough so that $n\delta^2/4f$ and $(0.5 - n\,\delta^2)/f$ are small. Additionally, to apply Model 10 the values of δ must be small, since for moderate to large values of δ the variance of $z(p)$ in Model 10 can be negative, even for very large degrees of freedom.

Summary

Seven approximate distributions have been identified for two functions of the obtained significance level. Asymptotically all provide reasonable results, but often they are applied to data based on small samples. Which approximations are most accurate in very small samples, and how large samples must be in order for each of the approximations to be used are the questions addressed in the investigations described below.

Small Sample Accuracy of the Asymptotic Theory

Analytical comparisons of exact and asymptotic distributions are

difficult when significance levels are of interest because of the complicated form of the cumulative distribution function of the sample significance value (the density was given above). Through the use of the computer, however, approximate and computed "exact" distributions can be compared.

Approximate cumulative distribution functions (based on the normal approximations outlined above) were compared to numerical evaluations of the exact cumulative distribution of the p, or equivalently of the noncentral t, for several sample sizes and effect sizes (δ values). Only the case of p's from one-sample t tests was studied. The generalization of these results to p's from other tests may not be justified.

The comparisons are based on the fact that if $Y = g(X)$ is a monotonic function of the random variable X, F is the cumulative distribution function (cdf) of X, and H is the cdf of Y, then for a particular value of $X = x$,

$$F(x) = H(y) = H(g(x)).$$

In the present case X represents the p value from the t statistic and $g(X)$ is the transformed p (i.e., $-\log(p)/n$ or $z(p) = -\Phi^{-1}(p)$). Theoretically the exact cdfs $F(p)$, $H(-\log(p)/n)$, and $H(-\Phi^{-1}(p))$ are equal because the logarithmic and inverse normal transformations are monotonic. By a similar argument the equivalence of $F(p)$ and the cdf for the t test can be shown.

Insert Table 1 about here

Table 1 presents the formulas for the expectations and variances of the approximate normal distributions of transformations of p's from one-

sample t statistics. The formulas for $c(\delta)$ and $\tau(\delta)$ are from Lambert and Hall (1983), and the others are derived from c and τ.

Computations

Comparisons of cdfs were made for effect sizes (δ values) ranging from 0.10 to 2.00 (for $n \leq 50$) or from 0.10 to 1.00 or 1.25 (for larger ns). Distributions based on larger sample and effect sizes were not examined because the ranges of p values for which $F(p)$ was not near either zero or one were quite extreme (i.e., the ps in those ranges were quite small) and did not represent interesting situations. Initial computations of $F(p)$ showed that even for relatively small effects ($\delta \approx 0.25$), $F(p)$ was changing (i.e., was not near zero or one) for very small values of p. For example, the median of the p distribution for $n = 10$ and $\delta = 0.25$ was .220, and for $n = 10$ and $\delta = 0.30$ it was .178.

For larger sample and effect sizes the distributions of ps covered even more extreme values of p. Because of this trend, ps associated with the central 90 percent of the range of the exact cdf of p were used as the domain for the distribution comparisons. Differences between the exact and approximate cdfs are largest in this range, since both distributions tend to 0 and 1 in the tails. Thus to some degree, the comparisons presented here based on computations across the middle range of ps overestimate differences between the exact and approximate distributions.

Insert Table 2 about here

Table 2 presents the p values used as limits in the 70 sample size by effect size patterns examined. These are the 5th and 95th percentiles of

the distribution of p for the given sample and effect sizes. Even the 95th percentiles of some of these distributions were quite small. The values were computed using IMSL (International Mathematical and Statistical Libraries, 1982) subroutines MDSTI and MDTN. The lower and upper limits for p produced lower and upper cdf values of .05 ± .002 and .95 ± .002, respectively.

Differences between the exact cdf and each approximation were calculated at 100 (or 101) points evenly spaced between the ps for the 5th and 95th percentiles of the exact cdf. Because the cdf of p rises rapidly (to values of .50 and higher) when δ is large, differences at closely spaced points within the first two of the 100 intervals were also examined.

The exact cdf of p was obtained as the exact cdf of a noncentral t variate. The central t for each p was first calculated with IMSL subroutine MDSTI. Then IMSL subroutine MDTN was used to get the probability of finding a more extreme value in the noncentral t distribution with $(n - 1)$ degrees of freedom and noncentrality parameter $\sqrt{n}\,\delta$.

Computational difficulties were encountered in obtaining the exact cdf of p for larger sample and effect sizes (i.e., when smaller ps and smaller intervals between ps were used). The algorithm used by the MDSTI subroutine to calculate the t variate for a given p obtained that t value by successive division of intervals in the t scale. Associated p values were calculated, and when the difference between the input p value and the calculated p was "small," the t associated with the calculated p was returned. MDSTI is a single precision subroutine, which is accurate to 14 significant digits on the Cyber 750 computer used for the computations. Thus differences smaller than 1.00×10^{-14} between input and calculated ps could not be

discriminated.

For larger sample sizes the domain of the exact cdf of p included ps smaller than 10^{-14} and was divided into even smaller intervals. Thus series of ps input to MDSTI (as many as 14) produced the same (output) t values, creating a step function for the ts and consequently for the computed "exact" distribution. Differences between the approximate cdf values (which were computed using IMSL subroutine MDNOR) and the computed "exact" function values were larger in absolute value near jumps in the step function. These larger differences suggest that the approximations are less accurate than they truly are.

Which ps were near the jumps could only be identified by using very closely spaced ps for computing the comparisons between cdfs. Thus it is not known whether most of the differences computed were for points near the jumps in the step function. Differences in the tables are reported to only two decimal places to portray this uncertainty.

Five measures of accuracy were obtained for each approximation in each sample size by effect size combination. The maximum absolute difference between the exact and the approximate cdf was obtained from either the widely spaced or closely spaced ps. The mean absolute difference, the square root of the mean squared difference, and the mean and variance of the signed differences were computed over the 100 or 101 widely spaced points.

Results for the Lognormal Approximation

The lognormal approximation was studied for four functions: Lambert's original normal approximation was used for $-\log(p)/n$, as well as for the modified functions $-\log(\sqrt{n}\ p)/n$, $-\log(\sqrt{2\pi n}\ p)/n$, and $-\log(2\sqrt{n}\ p)/n$. Tables 3 and 4 present the maximum absolute differences and the mean

absolute and signed differences for cdfs associated with these four statistics for several sample and effect size combinations.

Insert Tables 3 and 4 about here

The results showed the normal approximation based on $-\log(p)/n$ to be grossly inadequate for small samples and effects, which corroborated Lambert's earlier findings. The maximum difference between this and the exact cdf was never less than .15 in the situations examined. The signed mean difference was always the exact negative of the absolute mean difference, reflecting the fact that this approximation consistently underestimated the cdf.

The three modifications improved the accuracy of the lognormal approximation somewhat. For small to moderate effect sizes $-\log(\sqrt{n}\ p)/n$ performed the best of the three modifications. Even so it was still not very accurate, diverging from the exact cdf by as much as .33 for an effect size of $\delta = 0.10$ and a sample of 100. For larger effects the lognormal approximations based on the other two modifications (i.e., $-\log(\sqrt{2\pi n}\ p)/n$ and $-\log(2\sqrt{n}\ p)/n$) were quite close to the computed exact cdf, with maximum differences ranging between .02 and .04 for effect sizes of 1.0 or greater.

These comparisons of cumulative distributions indicated that although the unmodified lognormal approximation to the distribution of p was not accurate even for large (but still reasonable) values of n and δ, modified approximations were adequate when sample sizes were moderate to large ($n \geq 30$) and when effect sizes were large (i.e., greater than about 0.50).

It is likely that the decreased accuracy of the lognormal approximations for small effect sizes relates to the fact that the

distribution of $-\log(p)/n$ is exponential rather than normal under the null hypothesis. The form of the distribution may become less normal as δ decreases. Simulation studies may provide information about that question.

Results for Normal Approximations based on z(p)

The accuracy of approximations based on $z(p)$ was investigated in the same fashion as the accuracy of the lognormal approximations. Significance values were transformed via $-\Phi^{-1}(p)$ and standardized using one of the three sets of means and standard deviations from Table 1.

Insert Tables 5 and 6 about here

Tables 5 and 6 present the maximum absolute differences and the mean absolute and signed differences for the comparisons of the exact (computed) cdf and the three approximate cdfs. The differences were much smaller in all cases than those for any of the four lognormal approximations.

Differences for the approximations suggested in Lambert (1978) and in the technical appendix to Rosenthal and Rubin (1979) were very similar for smaller effect-size values, and both sets of differences were small. (See Table 5.) Because differences between the exact and approximate cdfs were so small for these approximations, results are only presented for $n \leq 50$.

As δ increased, the differences between the exact cdf and that based on the approximation from Rosenthal and Rubin increased. Rosenthal and Rubin commented that terms in the mean and variance involving δ need to be small for the standard normal approximation to apply; those terms also must be small for this approximation (from which the standard normal approximation was derived) to apply. Furthermore, when δ is larger than $[1- (2n)^{-1}]^{1/2}$

(which approaches one as $n \rightarrow \infty$), the variance of the Rosenthal and Rubin approximation is negative. Thus this approximation can not be used for larger δ values.

The Lambert approximation performed well on average for effect sizes larger than one, even when the sample size was small. Maximum absolute differences based on this approximation increased as δ increased, and were largest for small samples. When the sample size was large (e.g., $n > 30$), no discrepancies larger than .05 were found between the exact cdf and the Lambert approximation.

Results for the study of the accuracy of the standard normal approximation are given in Table 6. For small effect sizes the differences between the exact cdf and that based on the standard normal approximation were all small. For $\delta = 0.10$, even the maximum differences were equal to zero, when rounded to two decimal places. For $\delta \geq 1.00$, the maximum difference was never less than 0.29 and the average difference was always greater than .06. The standard normal approximation showed reasonable accuracy only for effect sizes less than about 0.40 or 0.50.

As both n and δ increased the accuracy of the approximation decreased. This finding is sensible in light of the fact that the nonnull distributions of ps, and thus of functions of ps such as $z(p)$, depend on both n and δ. The true distribution of p becomes much less variable as n increases, but the standard normal approximation does not account for that trend.

Summary and Recommendations

Findings

Asymptotic distributions based on $z(p)$ were reasonably accurate for most situations examined, whereas the lognormal approximations were

generally inaccurate. When the effect size was small the three approximations based on z(p) seemed equally accurate. For larger effect sizes, however, only the distribution derived by Lambert appeared useful.

The standard normal approximation was grossly inaccurate for $\delta > 0.75$, and the variance of the Rosenthal and Rubin approximation was usually computed to be negative for effect sizes of approximately 1.0 or greater.

Further, the two approximations provided by Rosenthal and Rubin were derived for cases in which ps arise from t tests, and may perform differently in other cases. Their performance under other conditions has not been examined. Alternatively, the distribution of z(p) given by Lambert may be applied with ps from many different (asymptotically normal) statistics simply by selecting the appropriate formulas for $c(\delta)$ and $\tau(\delta)$.

Implications for Suggested Uses of Significance Values

These results indicate that nothing would be gained by using log-transformed ps rather than z(p) values in the between-study comparisons suggested by Rosenthal and Rubin (1979) and Strube (1985). However, the results additionally indicate that caution must be observed when the standard normal approximation is applied. The distribution is somewhat inaccurate for moderate to large effects and becomes increasingly worse as sample sizes increase. These inaccuracies could be compounded in cumulative distributions for linear combinations of z(p) values.

In sum, Lambert's distribution is quite accurate and can be applied to ps not based on t tests, though considerable simplicity is gained (at the cost of little loss in accuracy) by using the standard normal approximation when the data arise from t tests and when small effects ($\delta \leq 0.30$) are expected.

References

Bahadur, R. R. (1960a). On the asymptotic efficiency of tests and estimates. Sankhya, 22, 229-252.

Bahadur, R. R. (1960b). Simultaneous comparison of the optimum and sign tests of a normal mean. In Olkin, I., Ghurye, S. G., Hoeffding, W., Madow, W. G., & Mann, H. B. (Eds.) Contributions to probability and statistics: Essays in honor of Harold Hotelling. Stanford: Stanford University Press.

Bahadur, R. R. (1967). Rates of convergence of estimates and test statistics. Annals of Mathematical Statistics, 38, 303-324.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand-McNally.

Cooper, H. M., & Rosenthal, R. (1980). Statistical versus traditional procedures for summarizing research findings. Psychological Bulletin, 87, 442-449.

Gibbons, J. D., & Pratt, J. W. (1975). P-values: Interpretation and methodology. The American Statistician, 29, 20-25.

Glass, G. V (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5, 3-8.

International Mathematical and Statistical Libraries, Inc. (1982). IMSL Library 1. (8th ed.). Houston, TX: Author.

Lambert, D. M. (1978). P-values: Asymptotics and robustness. Unpublished dissertation, University of Rochester.

Lambert, D. M., & Hall, W. J. (1982). Asymptotic lognormality of p-values. The Annals of Statistics, 10, 44-64.

Lambert, D. M., & Hall, W. J. (1983). Correction to "Asymptotic lognormality of p-values." The Annals of Statistics, 11, 348.

Pearson, E. S. (1938). The probability intergal transformation for testing goodness of fit and combining independent tests of significance. Biometrika, 30, 134-148.

Rosenthal, R. (1978). Combining results of independent studies. Psychological Bulletin, 85, 185-193.

Rosenthal, R. (1983). Assessing the statistical and social importance of the effects of psychotherapy. Journal of Consulting and Clinical Psychology, 51, 4-13.

Rosenthal, R. (1984). Meta-analytic procedures for social research. Beverly Hills: Sage.

Rosenthal, R., & Rubin, D. B. (1979). Comparing significance levels of independent studies. Psychological Bulletin, 86, 1165-1168.

Strube, M. J (1985). Combining and comparing significance levels from nonindependent hypothesis tests. Psychological Bulletin, 97, 334-341.

Table 1

Formulas for Means and Variances for the Approximate Normal Distributions[1]

Approximation	Mean	Variance
Lognormal	$\sqrt{n/2 \log(1 + \delta^2)}$	$\delta^2(1 + \delta^2/2)/(1 + \delta^2)$
Lambert's z	$\sqrt{n \log(1 + \delta^2)}$	$\delta^2(1+\delta^2/2)/[(1+\delta^2)\log(1+\delta^2)]$
Rosenthal & Rubin	$\sqrt{n}\,\delta\,[1 - n\delta^2/(4n-4)]$	$1 + (1/2 - n\,\delta^2)/(n - 1)$
Standard normal	$\sqrt{n}\,\delta$	1.0

1. The value n is the sample size on which p is based, and δ is the proposed alternate-hypothesis value of the effect size for the one-sample t.

Table 2

Fifth and 95th Percentiles of Distribution of p used as Limits in Computation

δ	Lower limit	Upper limit	Lower limit	Upper limit
	$n = 10$		$n = 25$	
0.10	2.66×10^{-2}	9.14×10^{-1}	1.72×10^{-2}	8.76×10^{-1}
0.25	1.01×10^{-2}	8.03×10^{-1}	2.57×10^{-3}	6.59×10^{-1}
0.30	6.99×10^{-3}	7.66×10^{-1}	1.30×10^{-3}	5.57×10^{-1}
0.40	3.53×10^{-3}	6.50×10^{-1}	3.35×10^{-4}	3.68×10^{-1}
0.50	1.89×10^{-3}	5.21×10^{-1}	7.55×10^{-5}	1.99×10^{-1}
0.75	3.70×10^{-4}	2.47×10^{-1}	1.90×10^{-6}	2.56×10^{-2}
1.00	8.36×10^{-5}	8.76×10^{-2}	5.48×10^{-8}	1.87×10^{-3}
1.25	2.23×10^{-5}	2.91×10^{-2}	1.80×10^{-9}	1.13×10^{-4}
1.50	7.11×10^{-6}	9.70×10^{-3}	8.76×10^{-11}	6.30×10^{-6}
1.75	2.18×10^{-6}	3.46×10^{-3}	5.29×10^{-12}	4.60×10^{-7}
2.00	8.28×10^{-7}	1.33×10^{-3}	3.80×10^{-13}	3.73×10^{-8}
	$n = 30$		$n = 40$	
0.10	1.56×10^{-2}	8.64×10^{-1}	1.25×10^{-2}	8.45×10^{-1}
0.25	1.67×10^{-3}	6.10×10^{-1}	8.99×10^{-4}	5.31×10^{-1}
0.30	8.44×10^{-4}	4.96×10^{-1}	3.41×10^{-4}	4.00×10^{-1}
0.40	1.64×10^{-4}	2.97×10^{-1}	4.44×10^{-5}	1.89×10^{-1}
0.50	2.83×10^{-5}	1.42×10^{-1}	4.82×10^{-6}	7.20×10^{-2}
0.75	3.92×10^{-7}	1.15×10^{-2}	1.71×10^{-8}	2.22×10^{-3}
1.00	5.01×10^{-9}	4.71×10^{-4}	6.52×10^{-11}	2.87×10^{-5}
1.25	1.00×10^{-10}	1.55×10^{-5}	3.18×10^{-13}	2.99×10^{-7}
1.50	2.55×10^{-12}	5.10×10^{-7}	2.39×10^{-15}	3.28×10^{-9}
1.75	8.46×10^{-14}	2.09×10^{-8}	2.39×10^{-17}	4.14×10^{-11}
2.00	3.51×10^{-15}	1.02×10^{-9}	3.78×10^{-19}	7.35×10^{-13}
	$n = 50$		$n = 75$	
0.10	9.38×10^{-3}	8.26×10^{-1}	6.25×10^{-3}	7.82×10^{-1}
0.25	4.56×10^{-4}	4.55×10^{-1}	1.11×10^{-4}	2.97×10^{-1}
0.30	1.46×10^{-4}	3.13×10^{-1}	1.98×10^{-5}	1.69×10^{-1}
0.40	1.18×10^{-5}	1.19×10^{-1}	6.59×10^{-7}	3.77×10^{-2}
0.50	8.50×10^{-7}	3.40×10^{-2}	1.48×10^{-8}	4.85×10^{-3}
0.75	8.56×10^{-10}	4.20×10^{-4}	6.56×10^{-13}	4.70×10^{-6}
1.00	8.61×10^{-13}	1.73×10^{-6}	2.49×10^{-17}	1.21×10^{-9}
1.25	1.17×10^{-15}	5.70×10^{-9}	1.30×10^{-21}	1.77×10^{-13}
1.50	2.57×10^{-18}	1.86×10^{-11}	-	-
1.75	8.15×10^{-21}	8.23×10^{-14}	-	-
2.00	4.12×10^{-23}	5.32×10^{-16}	-	-
	$n = 100$			
0.10	4.30×10^{-3}	7.38×10^{-1}		
0.25	2.70×10^{-5}	2.00×10^{-1}		
0.30	3.51×10^{-6}	8.92×10^{-2}		
0.40	4.19×10^{-8}	1.08×10^{-2}		
0.50	3.51×10^{-10}	6.55×10^{-4}		
0.75	6.23×10^{-16}	4.93×10^{-8}		
1.00	8.64×10^{-22}	6.35×10^{-13}		

Table 3

Accuracy of Lognormal Approximations to the Cumulative Distribution of p: $-\log(p)/n$ and $-\log(\sqrt{n}\,p)/n$

		$-\log(p)/n$			$-\log(\sqrt{n}\,p)/n$		
n	δ	Maximum absolute difference	Mean absolute difference	Mean difference	Maximum absolute difference	Mean absolute difference	Mean difference
10	0.10	.65	.47	-.47	.33	.19	.11
	0.25	.50	.42	-.42	.18	.06	-.01
	0.30	.47	.40	-.40	.17	.04	-.03
	0.40	.43	.38	-.38	.16	.06	-.06
	0.50	.41	.36	-.36	.15	.08	-.08
	0.75	.38	.32	-.32	.14	.10	-.10
	1.00	.36	.30	-.30	.14	.11	-.11
	1.25	.35	.29	-.29	.14	.11	-.11
	1.50	.34	.28	-.28	.14	.10	-.10
	1.75	.33	.27	-.27	.13	.10	-.10
	2.00	.33	.27	-.27	.12	.09	-.09
25	0.10	.57	.45	-.45	.37	.21	.19
	0.25	.42	.37	-.37	.10	.05	.03
	0.30	.39	.35	-.35	.10	.02	.01
	0.40	.36	.31	-.31	.09	.03	-.03
	0.50	.33	.28	-.28	.08	.04	-.04
	0.75	.30	.24	-.24	.08	.06	-.06
	1.00	.29	.21	-.21	.09	.06	-.06
	1.25	.28	.20	-.20	.09	.06	-.06
	1.50	.27	.19	-.19	.08	.06	-.06
	1.75	.26	.18	-.18	.08	.05	-.05
	2.00	.26	.18	-.18	.08	.05	-.05
30	0.10	.55	.44	-.44	.37	.21	.19
	0.25	.40	.36	-.36	.09	.04	.04
	0.30	.38	.34	-.34	.09	.02	.01
	0.40	.34	.30	-.30	.08	.02	-.02
	0.50	.32	.27	-.27	.08	.04	-.04
	0.75	.29	.22	-.22	.07	.05	-.05
	1.00	.27	.20	-.20	.08	.05	-.05
	1.25	.26	.18	-.18	.08	.05	-.05
	1.50	.26	.18	-.18	.08	.05	-.05
	1.75	.25	.17	-.17	.07	.05	-.05
	2.00	.24	.17	-.17	.07	.05	-.05

Table 3 (Continued)

Accuracy of Lognormal Approximations to the Cumulative Distribution of p: $-\log(p)/n$ and $-\log(\sqrt{n}\,p)/n$

		$-\log(p)/n$			$-\log(\sqrt{n}\,p)/n$		
n	δ	Maximum absolute difference	Mean absolute difference	Mean difference	Maximum absolute difference	Mean absolute difference	Mean difference
50	0.10	.50	.42	-.42	.37	.21	.20
	0.25	.36	.32	-.32	.08	.04	.03
	0.30	.34	.30	-.30	.06	.02	.01
	0.40	.30	.25	-.25	.06	.02	-.02
	0.50	.28	.22	-.25	.06	.03	-.03
	0.75	.25	.17	-.17	.06	.04	-.04
	1.00	.24	.15	-.15	.06	.04	-.04
	1.25	.23	.14	-.14	.06	.04	-.04
	1.50	.22	.13	-.13	.06	.04	-.04
	1.75	.22	.13	-.13	.06	.03	-.03
	2.00	.21	.12	-.12	.05	.03	-.03
100	0.10	.44	.39	-.39	.33	.19	.19
	0.25	.31	.26	-.26	.07	.02	.02
	0.30	.28	.24	-.24	.04	.01	-.00
	0.40	.25	.19	-.19	.03	.02	-.02
	0.50	.23	.16	-.16	.03	.02	-.02
	0.75	.20	.12	-.12	.03	.03	-.03
	1.00	.19	.11	-.11	.04	.03	-.03

Table 4

Accuracy of Lognormal Approximations to the Cumulative Distribution of p: $-\log(2\sqrt{n}\,p)/n$ and $-\log(\sqrt{2\pi n}\,p)/n$

		$-\log(2\sqrt{n}\,p)/n$			$-\log(\sqrt{2\pi n}\,p)/n$		
n	δ	Maximum absolute difference	Mean absolute difference	Mean difference	Maximum absolute difference	Mean absolute difference	Mean difference
10	0.10	.59	.37	.28	.65	.33	.32
	0.25	.30	.18	.17	.37	.21	.21
	0.30	.24	.14	.13	.31	.18	.17
	0.40	.15	.09	.08	.22	.12	.12
	0.50	.10	.05	.05	.16	.09	.08
	0.75	.05	.01	.01	.08	.03	.03
	1.00	.03	.01	-.01	.05	.02	.02
	1.25	.04	.02	-.02	.04	.01	.01
	1.50	.03	.02	-.02	.04	.01	.00
	1.75	.03	.02	-.02	.04	.01	.01
	2.00	.02	.01	-.01	.05	.01	.01
25	0.10	.59	.30	.30	.64	.33	.32
	0.25	.26	.15	.14	.32	.17	.17
	0.30	.20	.11	.11	.25	.13	.13
	0.40	.13	.06	.06	.17	.08	.08
	0.50	.08	.03	.03	.12	.05	.05
	0.75	.03	.00	.00	.06	.02	.02
	1.00	.02	.01	-.01	.03	.01	.00
	1.25	.02	.01	-.01	.03	.01	-.00
	1.50	.02	.01	-.01	.02	.01	-.00
	1.75	.02	.01	-.01	.03	.01	-.00
	2.00	.01	.01	-.01	.03	.00	.00
30	0.10	.58	.30	.30	.64	.32	.32
	0.25	.25	.14	.14	.30	.16	.16
	0.30	.19	.10	.10	.24	.12	.12
	0.40	.12	.05	.05	.16	.07	.07
	0.50	.08	.03	.03	.11	.04	.04
	0.75	.03	.01	-.01	.05	.01	.01
	1.00	.02	.01	-.01	.03	.01	.00
	1.25	.02	.01	-.01	.02	.00	-.00
	1.50	.02	.01	-.01	.02	.00	-.00
	1.75	.01	.01	-.01	.02	.00	-.00
	2.00	.01	.01	-.01	.03	.00	-.00

Table 4 (Continued)

Accuracy of Lognormal Approximations to the Cumulative Distribution of p: $-\log(2\sqrt{n}\,p)/n$ and $-\log(\sqrt{2\pi n}\,p)/n$

		$-\log(2\sqrt{n}\,p)/n$			$-\log(\sqrt{2\pi n}\,p)/n$		
n	δ	Maximum absolute difference	Mean absolute difference	Mean difference	Maximum absolute difference	Mean absolute difference	Mean difference
50	0.10	.55	.29	.29	.61	.31	.30
	0.25	.22	.11	.11	.26	.13	.12
	0.30	.16	.18	.08	.20	.09	.09
	0.40	.10	.04	.04	.13	.05	.05
	0.50	.06	.02	.02	.19	.03	.03
	0.75	.02	.01	.01	.04	.01	.01
	1.00	.01	.01	-.00	.03	.00	-.00
	1.25	.01	.01	-.01	.02	.00	-.00
	1.50	.01	.01	-.01	.02	.00	-.00
	1.75	.02	.01	-.01	.02	.00	-.00
	2.00	.01	.01	-.01	.02	.00	-.00
100	0.10	.50	.25	.25	.54	.27	.27
	0.25	.17	.07	.07	.20	.08	.08
	0.30	.13	.04	.04	.16	.05	.05
	0.40	.08	.02	.02	.10	.03	.03
	0.50	.05	.01	.00	.07	.01	.01
	0.75	.02	.01	-.01	.03	.00	.00
	1.00	.01	.01	-.01	.02	.00	-.00

Table 5

Accuracy of Two Approximations to the Cumulative Distribution of p: Lambert's and Rosenthal and Rubin's Approximations using z(p)

		Lambert			Rosenthal & Rubin		
n	δ	Maximum absolute difference	Mean absolute difference	Mean difference	Maximum absolute difference	Mean absolute difference	Mean difference
10	0.10	.00	.00	.00	.01	.00	.00
	0.25	.01	.00	.00	.01	.00	.00
	0.30	.01	.00	.00	.01	.00	.00
	0.40	.01	.01	.01	.01	.00	.00
	0.50	.01	.01	.01	.01	.01	.00
	0.75	.02	.01	.01	.08	.03	.01
	1.00	.03	.02	.02	-	-	-
	1.25	.04	.02	.02	-	-	-
	1.50	.06	.03	.03	-	-	-
	1.75	.08	.03	.03	-	-	-
	2.00	.10	.04	.04	-	-	-
25	0.10	.00	.00	.00	.00	.00	.00
	0.25	.00	.00	.00	.01	.00	.00
	0.30	.00	.00	.00	.00	.00	.00
	0.40	.01	.00	.00	.00	.00	.00
	0.50	.01	.00	.00	.01	.01	.00
	0.75	.01	.01	.01	.09	.02	.01
	1.00	.02	.01	.01	-	-	-
	1.25	.03	.01	.01	-	-	-
	1.50	.04	.01	.01	-	-	-
	1.75	.05	.02	.02	-	-	-
	2.00	.06	.02	.02	-	-	-
50	0.10	.00	.00	.00	.00	.00	.00
	0.25	.00	.00	.00	.00	.00	.00
	0.30	.00	.00	.00	.00	.00	.00
	0.40	.00	.00	.00	.00	.00	.00
	0.50	.00	.00	.00	.01	.01	.00
	0.75	.01	.00	.00	.07	.01	.01
	1.00	.01	.01	.01	-	-	-

Note. Dashes indicate that negative variances were computed for the approximation derived from Rosenthal and Rubin (1979).

Table 6

Accuracy of the Standard Normal Approximation to the Cumulative Distribution of z(p)

n	δ	Standard normal: Maximum absolute difference	Standard normal: Mean absolute difference	Standard normal: Mean difference
10	0.10	.00	.00	.00
	0.25	.02	.00	.00
	0.30	.02	.01	.01
	0.40	.04	.01	.01
	0.50	.06	.01	.01
	0.75	.15	.04	.04
	1.00	.29	.07	.07
	1.25	.45	.11	.11
	1.50	.60	.14	.14
	1.75	.78	.15	.15
	2.00	.89	.16	.16
25	0.10	.00	.00	.00
	0.25	.02	.00	.00
	0.30	.02	.00	.00
	0.40	.04	.01	.01
	0.50	.08	.02	.02
	0.75	.18	.04	.04
	1.00	.33	.08	.08
	1.25	.59	.10	.10
	1.50	.82	.11	.11
	1.75	.92	.11	.11
	2.00	.95	.11	.11
30	0.10	.00	.00	.00
	0.25	.02	.00	.00
	0.30	.02	.00	.00
	0.40	.05	.01	.01
	0.50	.08	.02	.02
	0.75	.18	.04	.04
	1.00	.35	.08	.08
	1.25	.64	.10	.10
	1.50	.85	.11	.11
	1.75	.94	.11	.11

Table 6 (Continued)

Accuracy of the Standard Normal Approximation to the Cumulative Distribution of z(p)

		Standard normal		
n	δ	Maximum absolute difference	Mean absolute difference	Mean difference
50	0.10	.00	.00	.00
	0.25	.02	.00	.00
	0.30	.03	.01	.01
	0.40	.05	.01	.01
	0.50	.08	.02	.02
	0.75	.19	.05	.05
	1.00	.46	.08	.08
	1.25	.77	.09	.09
100	0.10	.00	.00	.00
	0.25	.02	.00	.00
	0.30	.03	.01	.01
	0.40	.05	.01	.01
	0.50	.08	.02	.02
	0.75	.26	.06	.06
	1.00	.64	.08	.08